

学校编码: 10384

分类号\_\_\_\_\_密级\_\_\_\_\_

学号: X2007230041

UDC

厦门大学

\_\_\_\_\_硕士\_\_\_\_\_学 位 论 文

## 经营分析系统中的客户流失分析

Analysis of Customer Loss in Business Analytical System

作 者 姓 名: 吕幸

指导教师姓名: 廖明宏 教授

专 业 名 称: 软件工程

论文提交日期:

论文答辩时间:

学位授予日期:

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

2009 年 月

## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为( )课题(组)的研究成果,获得( )课题(组)经费或实验室的资助,在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

## 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（        ） 1. 经厦门大学保密委员会审查核定的保密学位论文，  
于        年        月        日解密，解密后适用上述授权。

（        ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年        月        日

---

## 摘要

电信市场竞争的加剧使得各电信企业客户的流失现象日益严重，建立高效的客户管理系统成为电信企业的必要选择。本论文以中国移动福建分公司开发的福建移动经营分析系统及移动客户流失分析系统为背景，阐述如何应用数据仓库及数据挖掘技术原理设计面向移动客户流失分析的数据仓库及数据挖掘模型，并使用大型数据库开发工具 `oracle 9i` 开发设计该数据仓库原型，使用拥有强大的报表制作和展示功能的数据展现工具 `Cognos` 动态交互展现客户流失多维数据，使用运用广泛、技术成熟、运行稳定的数据挖掘软件 `Clementine` 开发设计客户流失数据挖掘模型。

本论文主要围绕移动客户流失数据仓库设计、客户流失数据挖掘分析建模、客户流失 `OLAP` 前端展示三个方面进行研究。

在数据仓库方面，本文研究了数据仓库概念原理及数据仓库开发方法，对数据仓库开发模型进行重点研究，并设计出以移动客户流失分析为主题的数据仓库星型开发模型。运用大型数据库开发工具 `Oracle` 建立数据仓库原型，实现星型模型中的维度表及事实表的立方拓扑结构。

在客户流失数据挖掘分析建模方面，在前面工作的基础上，把处理完的训练数据输入数据挖掘软件工具，建立移动客户流失预测模型，并从准确率、覆盖率、提升率三方面对模型进行评估。

在 `OLAP` 数据展现方面，通过 `OLAP` 工具对数据进行综合、运算等处理，并通过建立不同的分析型应用，对数据仓库中的数据进行快速的、多角度的、复杂的分析，并为移动公司的各级管理人员提供灵活易用的用户界面。

最后，总结了论文的主要工作，阐述了进一步研究方向。

关键词：数据仓库；数据挖掘；经营分析系统；客户流失

---

## Abstract

Increased competition in the telecommunications market of the telecommunications companies make the loss of customers is becoming increasingly serious, the establishment of efficient customer management system has become necessary to choose the telecommunications business. In this paper, a China Mobile subsidiary Fujian Mobile, Fujian developed analysis system operators and mobile customers against the backdrop of the loss analysis system on the application of data warehouse and data mining technology for mobile customers design principle of the loss of the analysis of data warehouse and data mining model, and the use of large-scale database development tool for oracle 9i data warehouse development and design of the prototype, the use of statements has a strong function of the production and display of data to show the dynamic interaction Cognos tools to show customers the loss of multi-dimensional data, the use of the use of a wide range of technology is mature, stable operation of the data mining software Design Clementine data mining model churn.

In this paper, the loss of major clients around the mobile data warehouse design, OLAP front-end display of customer loss, customer loss of data mining analysis to study three aspects of modeling.

In data warehouses, this paper, the concept of the principle of data warehouse and data warehouse development methodology, the data warehouse focuses on the development of models and design analysis of a mobile client as its theme the fact that the loss of the data warehouse development model star. The use of large-scale database development tool oracle data warehouse to establish a prototype model to achieve the star of the dimension table and fact table in cubic topology.

The loss of data mining in customer analysis modeling, in front of the basis of the work, the processing of training data input software tool

---

of data mining, the establishment of mobile customer churn prediction model, and accuracy, coverage, enhance the rate of the three face models for evaluation.

In the OLAP data to show, through OLAP tools for data integration, processing, etc., and through the establishment of an analysis of the different applications of data warehouse data in a fast, multi-angle, complex analysis, and for mobile company managers at all levels to provide a flexible easy-to-use user interface.

Finally, the paper summarizes the major work on the direction of further research.

Key words: data warehouse; data mining; business analysis system; customer loss

# 目录

<b>第 1 章 绪论</b>	<b>1</b>
1.1 电信行业目前面临的市场挑战	1
1.2 电信行业实施客户流失分析的必要性	1
1.3 目前国内外研究的现状	2
1.3.1 国外研究的现状	2
1.3.2 国内的研究现状	3
<b>第 2 章 关键技术介绍</b>	<b>5</b>
2.1 数据仓库基本原理及相关概念	5
2.1.1 数据仓库简介	5
2.1.2 数据仓库主要概念	6
2.1.3 数据仓库与传统数据库的区别	11
2.2 数据挖掘基本原理及相关概念	13
2.2.1 数据挖掘的定义	13
2.2.2 数据挖掘的产生和发展	13
2.2.3 数据挖掘的任务	15
2.2.4 电信客户流失分析有关的数据挖掘方法	17
2.2.5 数据挖掘在电信行业客户关系管理中的应用	19
2.3 本章小结	20
<b>第 3 章 客户流失主题的数据仓库设计</b>	<b>21</b>
3.1 数据仓库模型的设计	21
3.1.1 概念模型的建立	21
3.1.2 逻辑模型的建立	22
3.1.3 物理模型的建立	25
3.2 数据抽取、转换和加载 (ETL)	27
3.2.1 数据抽取	27
3.2.2 数据的清洗和转换	29
3.2.3 数据的加载	30
3.3 本章小结	30
<b>第 4 章 经营分析系统中的客户流失分析模型</b>	<b>31</b>
4.1 商业理解	31
4.1.1 商业目标	31
4.1.2 挖掘问题定义	31
4.1.3 数据挖掘目标	32
4.2 数据准备	32
4.2.1 确定时间窗口	32
4.2.2 确定数据来源及预处理	33
4.3 客户流失分析建模	34
4.3.1 挖掘算法与挖掘软件的选择	34

---

4.3.2 建立客户流失预测模型 .....	36
4.4 客户流失分析模型的评估 .....	39
4.4.1 测试数据集的选择 .....	39
4.4.2 模型评价指标 .....	40
4.4.3 本文中的模型评价结果 .....	42
4.5 客户流失分析模型的运行结果 .....	43
4.6 本章小结 .....	44
<b>第 5 章 客户流失的 OLAP 前端展示 .....</b>	<b>45</b>
5.1 福建移动经营分析系统的总体设计 .....	45
5.2 经营分析系统中前端展示工具的选型: cognos .....	46
5.3 OLAP 的实现过程 .....	47
5.4 本章小结 .....	50
<b>第 6 章 总结与展望 .....</b>	<b>51</b>
<b>参考文献 .....</b>	<b>53</b>
<b>致谢 .....</b>	<b>55</b>



---

## Contents

Chapter 1	Introduction .....	1
1.1	The telecommunications industry's current market challenges.....	1
1.2	The need for analysis of customer churn .....	1
1.3	Research at home and abroad.....	2
1.3.1	Research abroad .....	2
1.3.2	Domestic Research.....	3
Chapter 2	Key Technology .....	5
2.1	The basic principles of data warehouse and related concepts.....	5
2.1.1	Introduction of Data Warehouse .....	5
2.1.2	The main concept of data warehouse .....	6
2.1.3	Data Warehouse and the difference between the traditional database ..	11
2.2	The basic principles of data mining and related concepts .....	13
2.2.1	The definition of data mining .....	13
2.2.2	The emergence of data mining and development .....	13
2.2.3	The task of data mining .....	15
2.2.4	Loss analysis of telecommunications customer data mining methods .	17
2.2.5	Data Mining in the Telecommunications Industry Customer Relationship Management Application.....	19
2.3	Summary of this chapter .....	20
Chapter 3	The loss of customer data warehouse design...	21
3.1	The design of data warehouse model.....	21
3.1.1	Concept model.....	21
3.1.2	Logic model.....	22
3.1.3	Physical model.....	25
3.2	Data extraction,Transformation and Load (ETL).....	27
3.2.1	Data extraction.....	27
3.2.2	Data cleansing and conversion .....	29
3.2.3	Loading data .....	30
3.3	Summary of this chapter .....	30
Chapter 4	Analysis System Model for Customer Churn.....	31
4.1	Business understanding.....	31
4.1.1	Business goals.....	31
4.1.2	The definition of mining.....	31
4.1.3	Data mining goals .....	32
4.2	Data preparation .....	32
4.2.1	Determine the time window .....	32
4.2.2	Identify data sources and preprocessing.....	33

---

4.3 Analysis of Customer Churn Modeling .....	34
4.3.1 Mining algorithm with the choice of mining software.....	34
4.3.2 Loss prediction model to build customer .....	36
4.4 Customer Churn Analysis Model of the Assessment.....	39
4.4.1 The choice of test data sets.....	39
4.4.2 Model evaluation .....	40
4.4.3 The results of the model to evaluate .....	42
4.5 Customer Churn Analysis Model of the results .....	43
4.6 Summary of this chapter .....	44
Chapter 5 Loss of customers to OLAP display .....	45
5.1 Analysis of the Fujian Mobile operating system design.....	45
5.2 Front-end display tools: cognos .....	46
5.3 The implementation of OLAP .....	47
5.4 Summary of this chapter .....	50
Chapter 6 Conclusion .....	51
References .....	53
Acknowledgment .....	55

## 第 1 章 绪论

### 1.1 电信行业目前面临的市场挑战

进入 21 世纪后,中国的移动通信服务市场竞争日渐激烈。电信、移动、联通、网通等运营商为了争夺客户,不断推出各种各样的资费政策,价格战在各处打响。而且营销的渠道也多种多样,各个代理商为了追求自己的利润,采取各种方法为用户办卡,很多用户不断换卡,以便叠加上更加优惠的资费,造成了资费的进一步复杂化。这些用户在各个运营商之间频繁流动,造成运营商运营成本居高不下<sup>[1]</sup>。

在这样的形式下,各运营商们这几年加强了对网络覆盖、支撑系统、营销网络和数据仓库应用的建设,不断增强自己的竞争实力,形成领先的希望扩大领先的优势,落后的奋起直追的竞争局面。

在我国,电信行业的信息化程度虽然是各个行业领先的,但是还是跟不上市场发展的程度。有的电信企业甚至还未做到实时出帐,用户的消费只有等到月底统一计算才能知道,跟不上时代的需求。国内电信行业急需走出目前“经验指导市场”的状况,真正实现从经验型决策到分析型决策的转变。

### 1.2 电信行业实施客户流失分析的必要性

根据 2-8 法则,80%的收益来自于 20%的客户贡献<sup>[2]</sup>。所以拥有一群忠诚的客户是电信行业得以发展的基础。

随着世界经济的全球化、市场的国际化,国内移动通信行业的市场已渐趋向合理且充分竞争。在电信企业面向市场,面向国内外众多竞争者努力创造更高价值的同时,客户流失不断增加,客户平均生命周期的不断缩短严重影响了企业的发展。在激烈的市场竞争和不断变化的市场需求面前,如何最大程度的降低客户流失率,传统的通过简单的查询方式、统计分析,或仅根据经验来了解客户行为。由于预测模型本身的复杂性及处理海量数据方面的局限性导致传统客户流失分析模型的有效性大打折扣。这些客户行为隐藏在诸如呼叫事务等庞大的数据集

中,数据挖掘起着极其重要的作用,数据挖掘技术是目前数据仓库领域最强大的分析手段,采用闭环方式控制,能把挖掘的结果反馈到决策过程中,基于严格数学计算的数据挖掘技术能彻底地改变以往电信企业在成功地获取客户以后无法监控客户的流失,无法实现客户关怀的状况。当前的电信行业需要把基于科学决策的客户关系管理全面引入到市场和营销工作中来。

电信行业是数据密集型行业,多年的业务支撑系统(BOSS)运行积累了大量的从客户资料到客户消费行为的数据。谁能正确地分析这些数据,得到有用的知识,谁就能更好地向用户提供服务,能够发现更多的商机,从而在竞争中获胜。电信企业必须保存用户的呼叫数据来计算费用,保存客户资料来提供客户服务,这些信息都是十分宝贵的。

用户流失前的行为有一定规律可循。通过利用数据挖掘技术,建立用户流失模型,定期产生可能流失的高危用户名单,并通过分析这些用户的消费行为特点,深层次地把握其需求,设计针对性的挽留政策。通过客户经理、外呼人员等进行一对一的客户关怀和挽留活动。由于挽留采取一对一的沟通方式,一方面避免了为控制流失而出台的优惠政策带来的大面积影响,另一方面,由于挽留的针对性比较强,也提高了挽留的成功率。

2003年,移动和联通的平均离网率均在10%以上,新增用户带来的收入一半被离网客户带来的损失抵消了。辛辛苦苦经营了一年得到这样的结果是相当让人失望的。所以在目前的电信行业建立一套科学的离网预警系统来辅助进行客户挽留工作是十分有必要的<sup>[3,4]</sup>。

## 1.3 目前国内外研究的现状

### 1.3.1 国外研究的现状

国外电信市场的竞争激烈程度超过了国内任何地方。降低运营成本,稳固自己的用户群是其十分注重的的问题。这些地方的电信运营商都已经建立起了自己的企业级的数据仓库,并在其上构建了完善的商业应用,尤其是客户流失分析的应用。不少知名的电信公司利用数据挖掘技术提升利润空间,在产品活动的生命周期中瞄准客户,确定变化的市场趋势并改善它在不同的市场划分中的洞察力,有

效提高了投资回报率。在客户流失分析方面存在不少成功的数据挖掘案例，尤其是竞争更为激烈的移动通信市场领域，例如，英国电信(BT)、美国的西南贝尔、Vanguard Cellular、AT&T Wireless 等公司通过建立流失客户的预测模型对客户流失进行管理和控制，对高流失概率的客户进行针对性的市场挽留工作，能够明显减小在网客户的流失率。因此，当前在北美和欧洲，对于客户流失行为预测技术的研究正逐渐的受到人们的关注<sup>[5]</sup>。

当前在北美和欧洲，由于其电信市场已经比较成熟，电信运营商的经营模式已经由传统的“技术驱动”过渡到了“市场驱动”和“客户驱动”的运营模式。国外整个数据仓库系统构建了很久，目前已经形成一个非常完善的架构体系。数据源的抽取到非常强壮的底层数据中心，以及在数据中心的基础上搭载的模型以及在相应的模型基础上开发应用。世界各地的许多运营商不仅建立了经营分析系统而且很多已经从中获得了巨大的收益<sup>[6]</sup>。

### 1.3.2 国内的研究现状

在国内，电信业的快速发展是近几年的事，这段时间内电信业的精力主要集中在抢占市场上，财务的手段也主要是用经过初步的市场调研和表面上的数据分析得出的结果来制定新的服务策略<sup>[7]</sup>。技术研究主要是业务运营支持系统(Business Operation Support System, BOSS)的更新换代，如对计费系统、客户服务系统、财务管理系统的开发和完善。这些系统中保留着大量的客户通话行为数据、账单、投诉咨询和客户基本信息数据。由于电信企业忙于业务扩张，所以无暇顾及对这些数据的挖掘分析。随着国内的各大移动运营商经营理念和运营模式的改变，必将对电信客户流失预测的研究起到巨大的推动作用。近一两年来，一些有实力的电信企业也开始考虑到挽留高价值客户的必要才逐步着手对这些历史数据进行挖掘。由于是刚刚起步，所以国内电信企业缺乏专业性的数据挖掘分析人员，对数据挖掘应用认识不够深入。数据挖掘的原型建模往往通过外部厂商实现，自身大多缺乏独立建模分析的能力。严重缺乏数据挖掘应用实施的经验以及相关的营销数据支撑。并且国内的大多电信运营商现有的面向事务的数据在质量，完整性和一致性上都存在许多问题，必须投入大量的精力去进行数据的抽取，净化和处理，并且当前国内未建立起完善的信用体系，导致现有客户价值模

型的有效性大大降低。由于起步的时间不长，大部分都只是试探性的建立简单的模型，有的还处于调研与可行性分析阶段，没有实际投入应用的产品<sup>[8]</sup>。

由此可见，国内电信企业在应用数据挖掘方面与国外同行相比还有很大的差距，在对数据挖掘技术的理解、营销理念、管理模式等方面都有待提高。

厦门大学博硕士论文摘要库

## 第 2 章 关键技术介绍

### 2.1 数据仓库基本原理及相关概念

随着计算机技术的迅速发展,信息处理技术也得到了长足的发展,从 70 年代中期的 MIS 系统发展到现代的数据仓库(Data Warehouse)技术。许多厂商都在开发自己的数据仓库解决方案,并通过各种渠道大力推广。对数据仓库最大的误解可能是把它当作一个现成的可以直接买来使用的产品<sup>[9]</sup>。事实上,数据仓库和数据库不同,它不是现成的软件或者硬件产品。比较确切地说,数据仓库是一种解决方案,是对原始的操作数据进行各种处理并转换成有用信息的处理过程,用户可以通过分析这些信息从而做出策略性的决策。因此,在很多场合,我们也把数据仓库系统称为决策支持系统。由于这个原因,数据仓库的用户不是类似银行柜员的终端操作人员,而是针对各个业务部门的用户和有关决策人员。因此,数据仓库的用户比传统的 OLTP (联机事务处理: On-line Transaction Processing) 用户少得多。

#### 2.1.1 数据仓库简介

数据仓库一个比较公认的定义:数据仓库(Data Warehouse)是一个面向主题的(Subject Oriented)、集成的(Integrate)、相对稳定的(Non-Volatile)、反映历史变化(Time Variant)的数据集合,用于支持管理决策<sup>[10]</sup>。

面向主题可理解是某个具体的业务。例如:主题可以是产品的销售情况,客户的财务状况,资产负债表,现金流量表等。面向主题说明数据仓库的数据集中了这些主题的所有相关数据,或者说数据仓库的数据是为这些主题而组织的(当然这之间包含数据抽取、清理、转换、装载的过程)。面向主题说明了数据仓库的业务特性,数据仓库关注着数据的业务关系和业务含义。相对于操作性数据,关注数据的逻辑关系和事务处理。例如:一个客户信息,对于操作性数据最重要的是这个客户的 ID,通过客户 ID 就可以找到相关信息,而对于客户的性别、年龄等或许根本就不影响 OLTP 的运行。而对于 OLAP,客户的性别、年龄却很重要,因为正是这些信息影响分析的结论。

**集成性** 集成则说明了数据仓库数据的多元性和一致性。为了获取某个主题全面的数据，数据仓库可能需要获取不同事务处理系统的数据，甚至是从文件、Internet 获取数据。然而这些数据不是拿过来就可以用的，这些异构的数据必须进行一致性处理(进行数据清理)。例如：以某产品的销售情况作为主题，产品的原材料、制造过程等数据在制成管理系统中，客户数据在客户管理系统中，销售倾向、销售成本在帐务管理系统中，而这些数据是相对独立，不一致的。集成就是要保证这些信息的完整和一致。

**相对稳定** 相对的对象是 OLTP，稳定指的更新的速度。说明数据仓库不是实时的(已经有实时数据仓库的提法，这个定义快过时了)。也就是说数据仓库要隔一定时间在去抽取数据，或者需要时才去抽取数据。另一层含义是指，数据被抽取之后就不被修改，将长期保存(完全的不做修改不大可能，很多 OLAP 都提供回写功能)。

**反应历史变化** 反应历史变化说明了数据仓库数据的连续性。大部分的操作型数据都有当前间、历史区别，他们只关心当前数据，对于历史数据将被封存、甚至被删除。而对于数据仓库数据是连续的不存在当前数据和历史数据的区别，只属于某个时间点。数据的变化轨迹能连续的展示。

**用于支持管理决策** 其实数据仓库并不仅仅用于 DSS，对不同的需求有着不同的用途。有的用数据仓库技术仅仅是为了提高数据查询性能。数据仓库的最大任务是：为 DSS 提供全面、一致、准确的数据，DSS 对这些数据进行加工处理以合理的方式展示给相关人员，把数据变成信息或者深加工形成知识。

### 2.1.2 数据仓库主要概念

#### (1) 分割

数据分割是数据仓库设计一项重要内容，是提高数据仓库性能的一项重要技术。数据的分割是指把逻辑上是统一整体的数据分割成较小的、可以独立管理的物理单元(称为分片)进行存储，以便于重构、重组和恢复，以提高创建索引和顺序扫描的效率。

数据的分割使数据仓库的开发人员和用户具有更大的灵活性。选择适当的数据分割的标准，一般要考虑以下几方面因素：数据量(而非记录行数)、数据分析



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库